

DATA ANALYTICS PORTFOLIO

Mrs. Dounia El Yousoufi



About Me



Data Analyst | Business Intelligence | Sustainability



Skills:

Data Analytics: Visualization | Statistical Analysis | Predictive Modelling

Tech & Tools: Python | SQL | Tableau | Excel | Machine Learning

Sustainability & Research: Data-Driven Sustainability Insights | Environmental Data Analysis | Research & Policy Evaluation



Education:

Data Analytics Certificate, *CareerFoundry*

MSc Sustainable Development, *Università degli Studi di Milano*

BSc Environmental Software Engineering, *Université Cadi Ayyad*



Languages:

Arabic | English | French | Italian (Int.) | German (Beg.)




01






PORTFOLIO SUMMARY



PORTFOLIO OVERVIEW

 **Objective**
Demonstrate
analytical expertise across
various fields.

 **Projects Covered**

 **Tools Used**
 Python |  SQL |
 Tableau |  Excel

◆ **Healthcare**
Influenza Season
Staffing Plan

◆ **Entertainment**
Rockbuster Stealth
Movie Analytics

◆ **E-Commerce**
Instacart Market
Basket Analysis

◆ **Finance**
Customer Churn
Analysis

◆ **Sustainability**
Climate Change
Impact on
Agriculture



PROJECT DETAILS

Project	Industry	Focus Area	Key Skills	Tools Used	Resources
Influenza Season US	Public Health	Medical Staff Planning	Data Cleaning, Integration, Hypothesis Testing, Visualisation	Excel, Tableau	Tableau
Rockbuster Stealth	Entertainment	Business Intelligence	SQL Querying, Filtering, Summarising, Joins, Subqueries, CTE	SQL, Tableau	GitHub , Tableau
Instacart	E-Commerce	Marketing Strategy	Data Wrangling, Merging, Grouping, Deriving Variables	Python, Excel	GitHub
Pig E. Bank	Banking & Finance	Customer Retention Strategy	Predictive Analysis, Decision Tree Modelling	Excel, PowerPoint	Excel
Climate Change & Agriculture	Sustainability	Impact of Climate on Crop Yield	Machine Learning (Regression, Clustering), Time Series Analysis, Visualization	Python	GitHub , Tableau



02

Influenza Season Staffing



Project Overview



Project Title

**Influenza Season Staffing
Analysis**



Objective

Optimise medical staffing allocation by analysing influenza mortality trends and vulnerable populations across U.S. states.



Business Context

📌 Flu season increases hospital demand, requiring strategic staffing.

📌 This project helps predict staffing needs across different states.



Data & Methodology



Data Sources

CDC – Influenza Mortality Data

U.S. Census Bureau – Population Demographics



Key Questions

- ✓ Which states have the highest vulnerable populations?
- ✓ Where is additional medical staff most needed?
- ✓ How does the 65+ population correlate with flu-related deaths?



Tools & Techniques

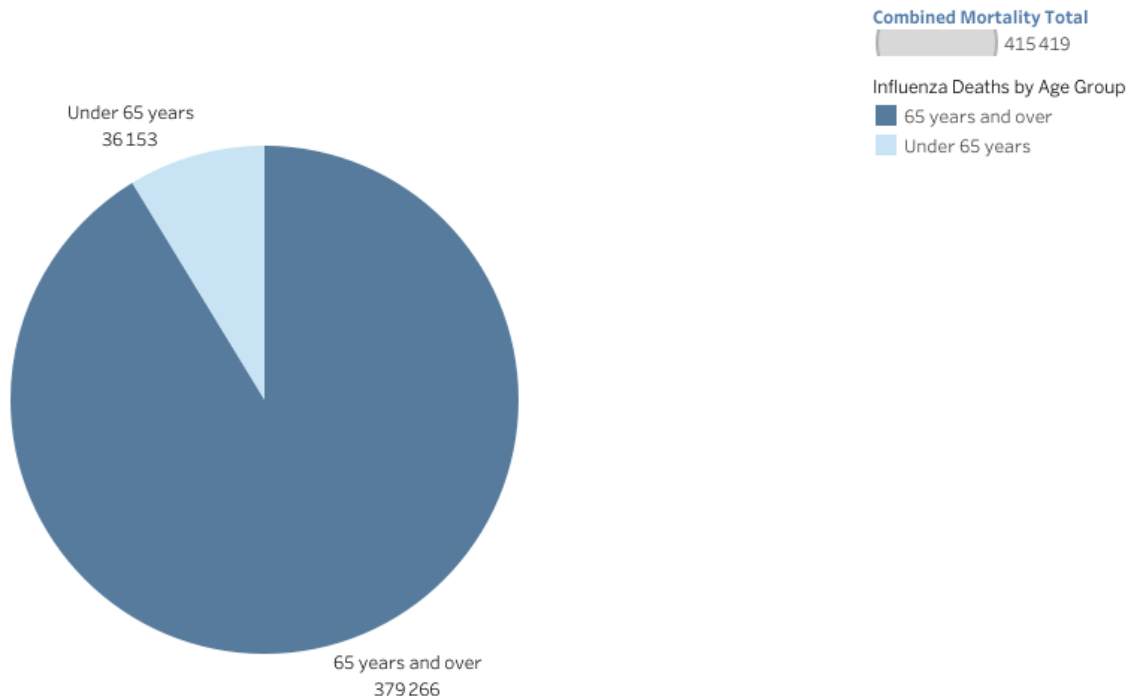
Data Cleaning & Transformation
(Excel)

Trend Analysis & Visualisation
(Tableau)



Influenza Mortality Breakdown (2009–2017)

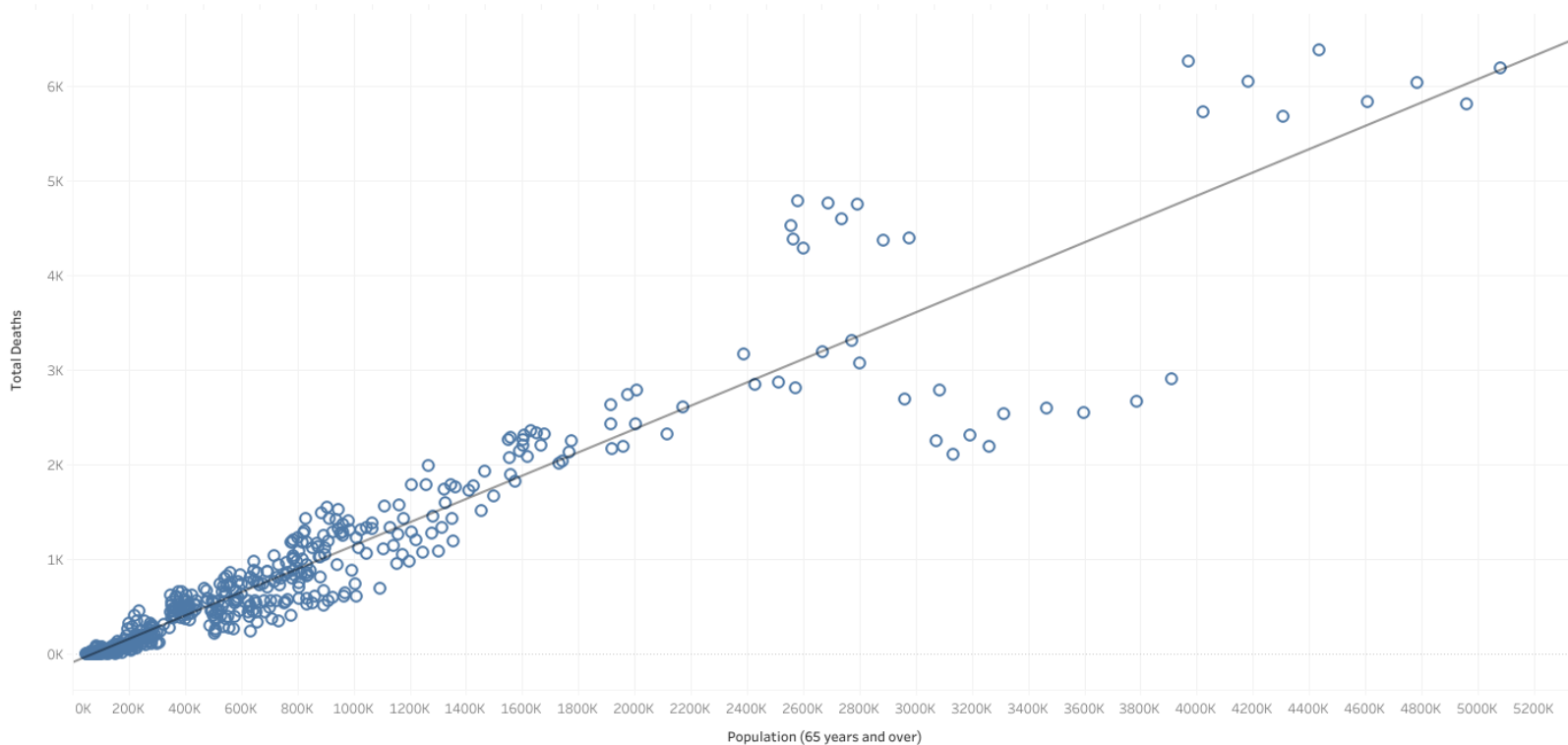
Influenza Mortality in the United States: 2009–2017



Insight: Shows overall influenza-related mortality distribution across different age groups.

Senior Population vs. Influenza-Related Deaths

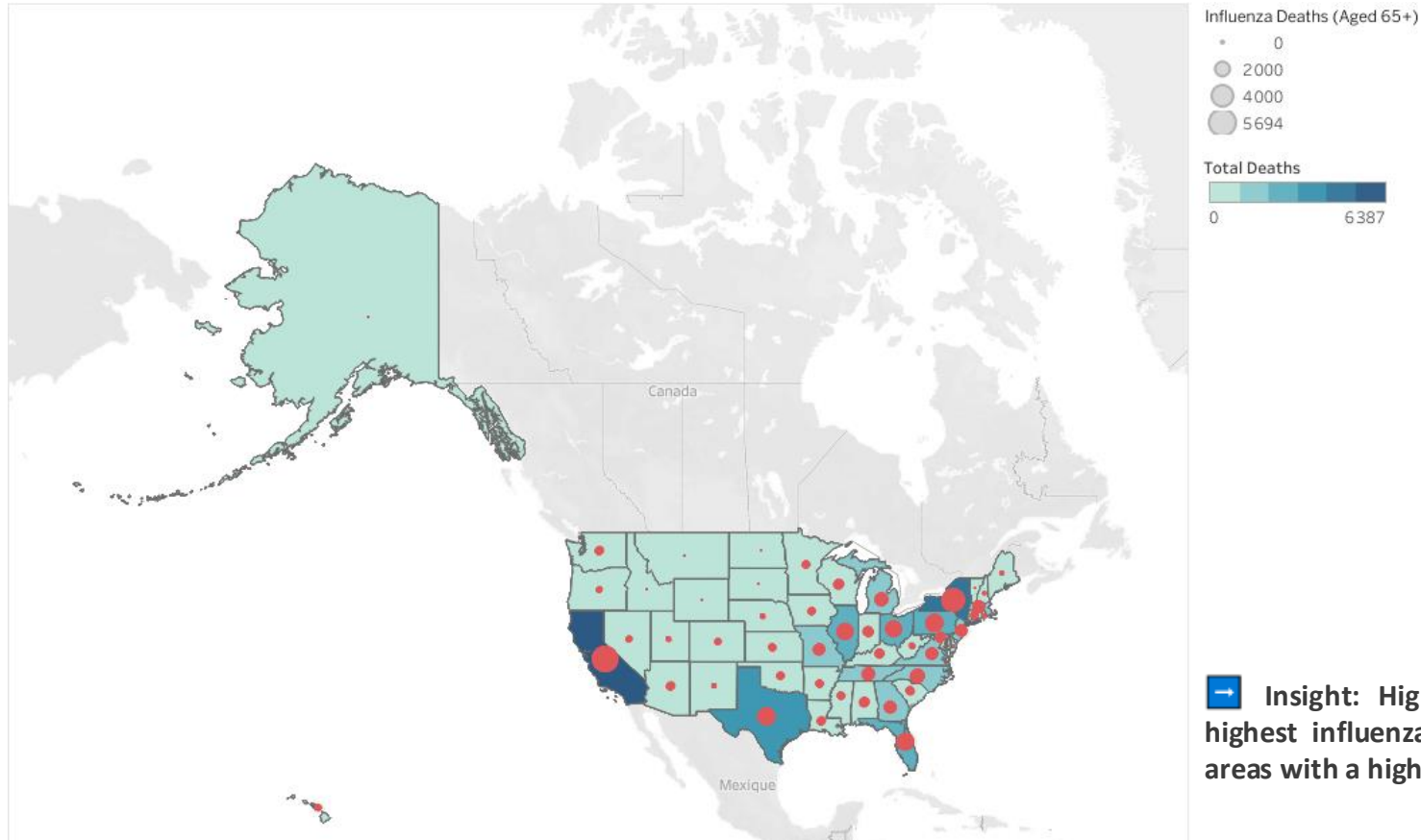
Relationship Between the Population Aged 65 and Over and Influenza-Related Deaths in the US (2009–2017)



Insight: Examines the correlation between the population aged 65+ and flu-related deaths to determine high-risk areas.

Where is Additional Medical Staff Needed?

Influenza Mortality Across the U.S. (2009–2017): Focus on Seniors Aged 65+



Insight: Highlights states with the highest influenza mortality, focusing on areas with a high senior population.

Summary of Findings and Actionable Recommendations

Key Findings:



High Mortality in Age 65+

Significant increase in influenza-related deaths among older adults.



High Mortality in Specific States

New York and similar states have the highest rates.



Annual Trends

Peak periods highlight need for focused resource allocation.

Recommendations :



Prioritize Staffing in High-Risk States

Focus additional staffing during peak flu periods.



Increase Vaccination Drives

Target vulnerable populations in high-risk states.



Update Staffing Plans Regularly

Use real-time flu data for better preparedness.



03

Rockbuster Stealth



Project Overview



Project Title

**Rockbuster Stealth LLC – Online
Platform Data Analysis**




Objective

Support Rockbuster's shift to online by analysing data to inform strategic decisions on customer behaviour, movie performance, and market trends.



Business Context

 Rockbuster is transitioning to an online service. Data analysis will drive decisions for customer engagement, revenue, and market expansion.



Data & Methodology



Data Sources

Rockbuster data set



Key Questions

- ✓ Top-performing and underperforming movies?
- ✓ Customer distribution & high-value segments?
- ✓ Rental duration & regional sales trends?



Tools & Techniques

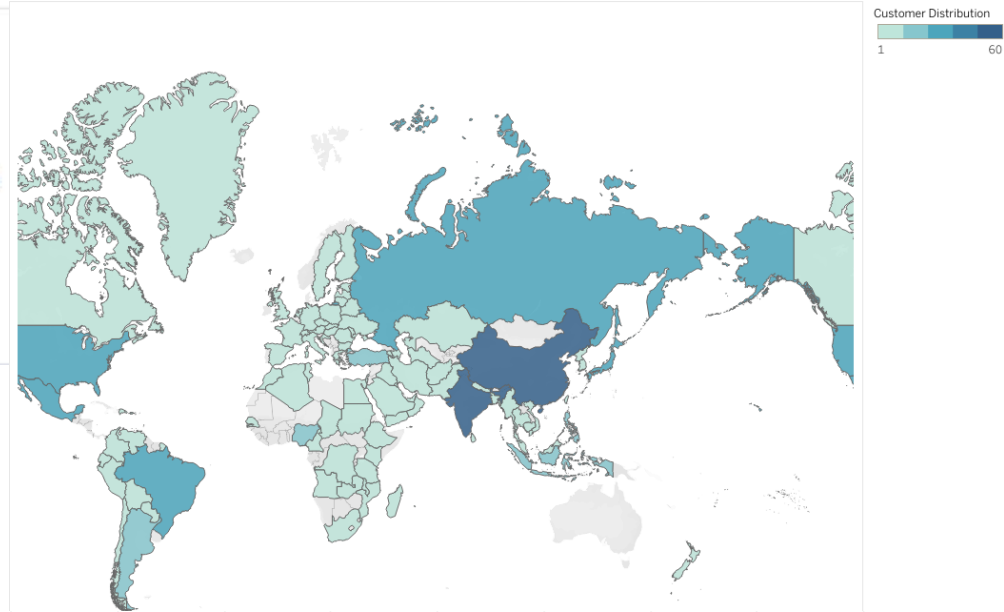
Data Cleaning & Transformation
(SQL)

Trend Analysis & Visualisation
(Tableau)



Top Customers

Customer Distribution by Country

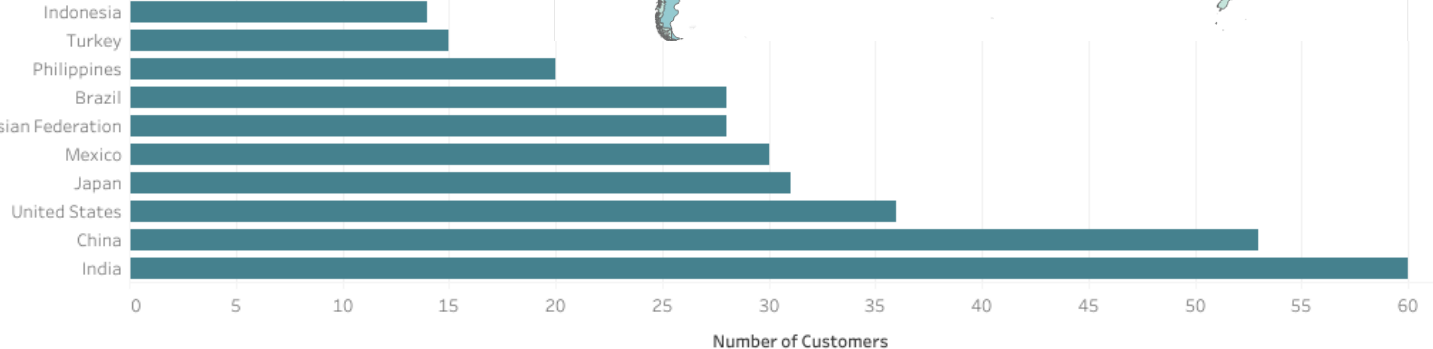


Query Query History

```
1 SELECT
2     country.country AS Country,
3     COUNT(customer.customer_id) AS Customer_Count
4 FROM
5     customer
6 INNER JOIN address ON customer.address_id = address.address_id
7 INNER JOIN city ON address.city_id = city.city_id
8 INNER JOIN country ON city.country_id = country.country_id
9 GROUP BY country.country
10 ORDER BY Customer_Count DESC
11 LIMIT 10;
```

	country character varying (50)	customer_count bigint
1	India	60
2	China	53
3	United States	36
4	Japan	31
5	Mexico	30
6	Brazil	28
7	Russian Federation	28
8	Philippines	20
9	Turkey	15
10	Indonesia	14

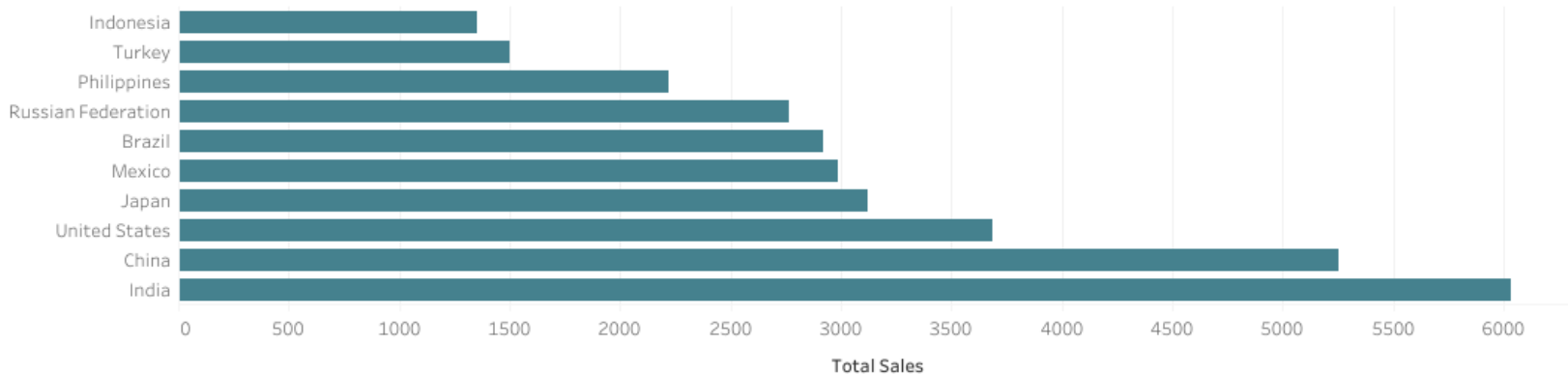
Top 10 Country by Customers



Top 10 Countries in Sales Distribution

Query	Query History		country character varying (50)	total_sales numeric
1	SELECT	1	India	6034.78
2	co.country,	2	China	5251.03
3	SUM(p.amount) AS total_sales	3	United States	3685.31
4	FROM payment p	4	Japan	3122.51
5	JOIN customer c ON p.customer_id = c.customer_id	5	Mexico	2984.82
6	JOIN address a ON c.address_id = a.address_id	6	Brazil	2919.19
7	JOIN city ci ON a.city_id = ci.city_id	7	Russian Federation	2765.62
8	JOIN country co ON ci.country_id = co.country_id	8	Philippines	2219.70
9	GROUP BY co.country	9	Turkey	1498.49
10	ORDER BY total_sales DESC	10	Indonesia	1352.69
11	LIMIT 10;			

Top 10 Countries in Sales Distribution



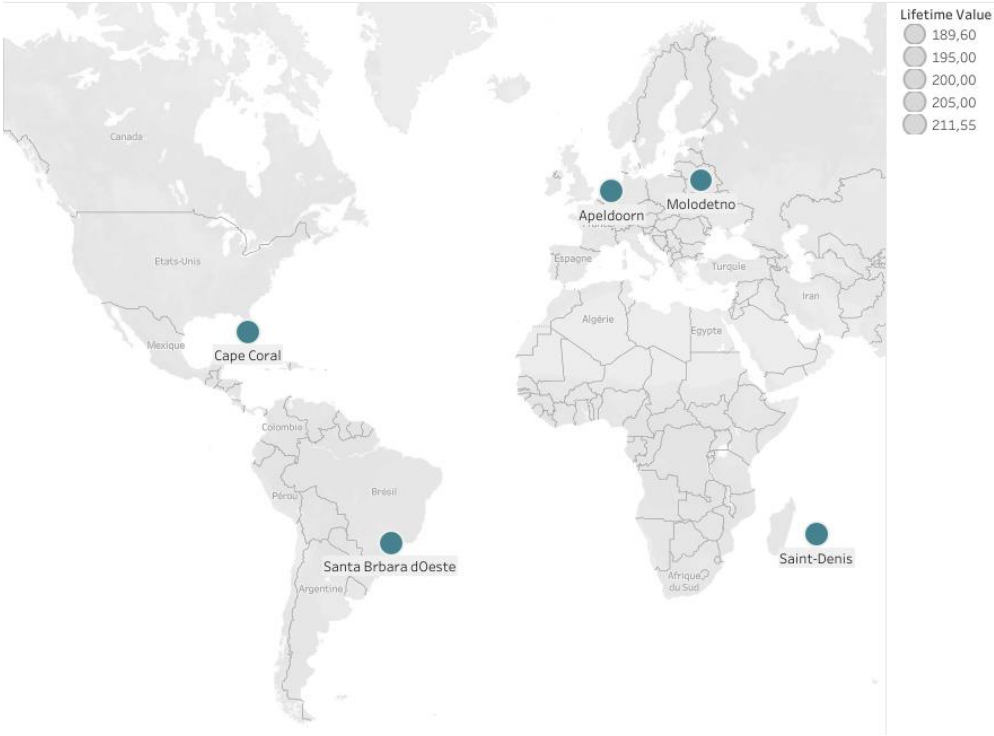
Customers with High Lifetime Value by Location

Query Query History

```
1 SELECT
2     co.country,
3     ci.city,
4     SUM(p.amount) AS lifetime_value
5 FROM customer c
6 JOIN address a ON c.address_id = a.address_id
7 JOIN city ci ON a.city_id = ci.city_id
8 JOIN country co ON ci.country_id = co.country_id
9 JOIN payment p ON c.customer_id = p.customer_id
10 GROUP BY co.country, ci.city
11 ORDER BY lifetime_value DESC
12 LIMIT 5;
```

	country character varying (50)	city character varying (50)	lifetime_value numeric
1	Runion	Saint-Denis	211.55
2	United States	Cape Coral	208.58
3	Brazil	Santa Brbara dOeste	194.61
4	Netherlands	Apeldoorn	191.62
5	Belarus	Molodetno	189.60

Top 5 Geographic Areas of High Lifetime Value Customers



Customer Overview

Query Query History

```
1 SELECT
2     store_id,
3     COUNT(store_id) AS customers_count
4 FROM customer
5 GROUP BY store_id
6 ORDER BY customers_count DESC;
```

	store_id smallint	customers_count bigint
1	1	326
2	2	273



599 Customers



584 Active
Customers



All Customers
Acquired on
14-Feb-2006

Query Query History

```
1 SELECT
2     MAX(customer_id) AS Total_number_customer,
3     MAX(store_id) AS Max_store_id,
4     MIN(store_id) AS Min_store_id,
5     MAX(create_date) AS latest_acquisition_date,
6     MIN(create_date) AS earliest_acquisition_date,
7     COUNT(active) AS active_customer
8 FROM customer
9 WHERE active > 0;
```

	total_number_customer integer	max_store_id smallint	min_store_id smallint	latest_acquisition_date date	earliest_acquisition_date date	active_customer bigint
1	599	2	1	2006-02-14	2006-02-14	584



326 Customers



273 Customers

Film Descriptive Statistics

Query Query History

```
1 SELECT
2     COUNT(film_id) AS Total_number_films,
3     MAX(rental_duration) AS Max_rental_duration,
4     MIN(rental_duration) AS Min_rental_duration,
5     AVG(rental_duration) AS Avg_rental_duration,
6     MAX(rental_rate) AS Max_rental_rate,
7     MIN(rental_rate) AS Min_rental_rate,
8     AVG(rental_rate) AS Avg_rental_rate,
9     MAX(length) AS Max_length,
10    MIN(length) AS Min_length,
11    AVG(length) AS AVG_length,
12    MAX(release_year) AS Max_year,
13    MIN(release_year) AS Min_year,
14    AVG(release_year) AS AVG_year
15 FROM film;
```

	total_number_films bigint	max_rental_duration smallint	min_rental_duration smallint	avg_rental_duration numeric	max_rental_rate numeric	min_rental_rate numeric	avg_rental_rate numeric
1	1000	7	3	4.9850000000000000	4.99	0.99	2.9800000000000000
max_length smallint	min_length smallint	avg_length numeric	max_year integer	min_year integer	avg_year numeric		
185	46	115.2720000000000000	2006	2006	2006.0000000000000000		

Rental Duration

Highest: 7 days

Lowest: 3 days

Average: 5 days

Rental Rate

Highest: \$4.99

Lowest: \$0.99

Average: \$2.98

Film Length

Highest: 118 min

Lowest: 46 min

Average: 115 min



All films released in 2006



Most Frequent Film Rating: PG-13

Query Query History

```
1 SELECT MODE() WITHIN GROUP (ORDER BY rating) AS Mode_rating
2 FROM film;
```

	mode_rating mpaa_rating
1	PG-13

Performance Analysis

Top 10 Performing Films

Query Query History

```
1 SELECT
2     f.title,
3     SUM(p.amount) AS revenue,
4     SUM(f.rental_duration) AS total_rental,
5     f.rating
6 FROM film f
7 JOIN inventory i ON f.film_id = i.film_id
8 JOIN rental r ON i.inventory_id = r.inventory_id
9 JOIN payment p ON r.rental_id = p.rental_id
10 GROUP BY f.film_id, f.title, f.rating
11 ORDER BY revenue DESC
12 LIMIT 10;
```

	title character varying (255) 🔒	revenue numeric 🔒	total_rental bigint 🔒	rating mpaa_rating 🔒
1	Telegraph Voyage	215.75	75	PG
2	Zorro Ark	199.72	84	NC-17
3	Wife Turn	198.73	81	NC-17
4	Innocent Usual	191.74	78	PG-13
5	Hustler Party	190.78	66	NC-17
6	Saturday Lambs	190.74	78	G
7	Titans Jerk	186.73	108	PG
8	Harry Idaho	177.73	135	PG-13
9	Torque Bound	169.76	69	G
10	Dogma Family	168.72	140	G

Worst 10 Performing Films


Query Query History


```
1 SELECT
2     f.title,
3     SUM(p.amount) AS revenue,
4     SUM(f.rental_duration) AS total_rental,
5     f.rating
6 FROM film f
7 JOIN inventory i ON f.film_id = i.film_id
8 JOIN rental r ON i.inventory_id = r.inventory_id
9 JOIN payment p ON r.rental_id = p.rental_id
10 GROUP BY f.film_id, f.title, f.rating
11 ORDER BY revenue ASC
12 LIMIT 10;
```


	title character varying (255) 🔒	revenue numeric 🔒	total_rental bigint 🔒	rating mpaa_rating 🔒
1	Oklahoma Jumanji	5.94	42	PG
2	Duffel Apocalypse	5.94	30	G
3	Texas Watch	5.94	42	NC-17
4	Freedom Cleopatra	5.95	25	PG-13
5	Young Language	6.93	42	G
6	Rebel Airport	6.93	49	G
7	Treatment Jekyll	6.94	18	PG
8	Cruelty Unforgiven	6.94	42	G
9	Lights Deer	7.93	49	R
10	Japanese Run	7.94	36	G


Summary of Findings and Actionable Recommendations


Key Findings:

 Top film: "Telegraph Voyage" (\$215.75), lowest: "Oklahoma Jumanji" (\$5.94)


 Customers split: Store 1 (326), Store 2 (273)


 Sales lead by India (\$6034.78), followed by China and the U.S.


 High lifetime value customers in Reunion, U.S., and Brazil


 Most Frequent Film Rating: PG-13


Recommendations:

 Target marketing in top-performing countries.

 Optimize inventory by reducing low-performing films.

 Expand catalog with high-demand genres.

 Implement loyalty programs for high-value customers.

 Explore growth in underperforming regions.



04

Instacart



Project Overview



Project Title

**Instacart Grocery Basket
Analysis**



Objective

To perform exploratory data analysis to uncover customer behaviour and sales patterns, and provide actionable insights for targeted marketing strategies.



Business Context

📌 Instacart aims to enhance ad targeting and sales by understanding customer behaviour, loyalty, and regional differences, boosting campaign relevance and sales in key areas.



Data & Methodology



Data Sources

Instacart Open-Source Data (via Kaggle)

Custom Customer Dataset
(provided by CareerFoundry)

Data Dictionary (Instacart)



Key Questions

- ✓ What are the busiest order days/hours?
When do customers spend the most?
- ✓ Which product types and price ranges are most popular?
- ✓ How do customer behaviours differ by loyalty, region, income, and family status?



Tools & Techniques

Data Cleaning & Transformation
(Python)

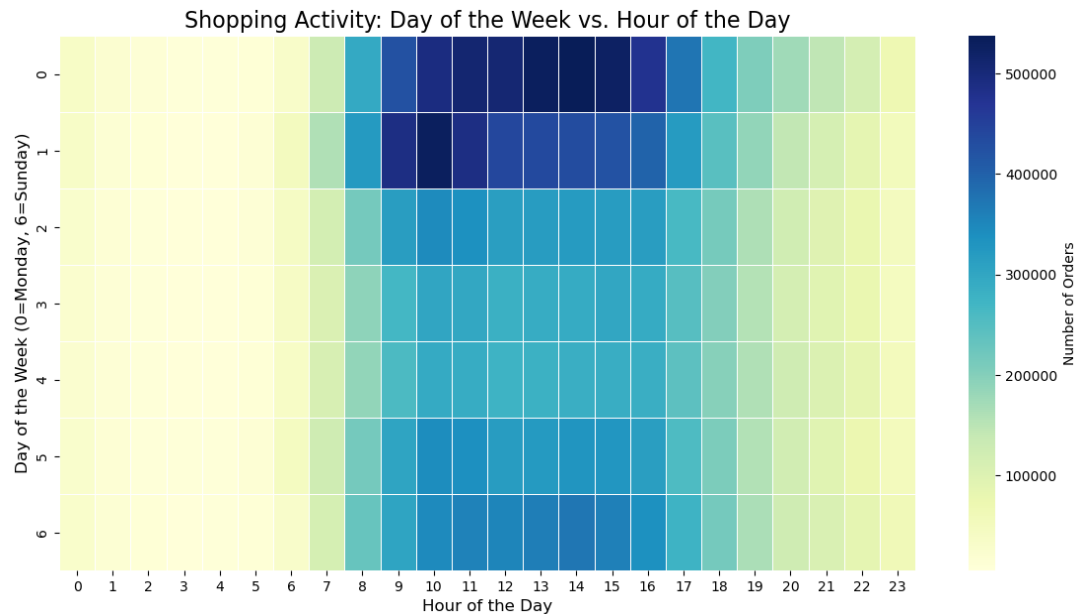
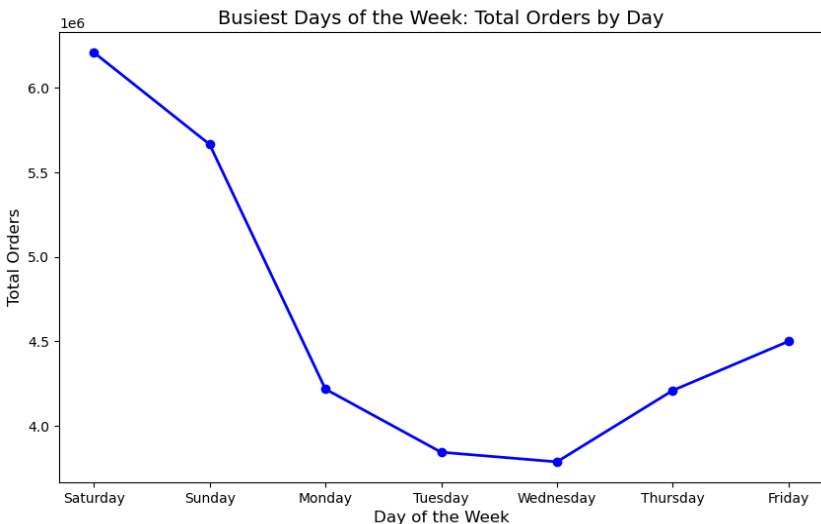
Trend Analysis & Visualisation
(Matplotlib, Seaborn)

Customer Segmentation & Profiling
(Pandas, NumPy)

Exploratory Data Analysis
(Jupyter Notebook)



Order Timing Insights for Ad Optimisation



Order Timing Insights

Busiest Day: Saturday

Least Busy Day: Wednesday

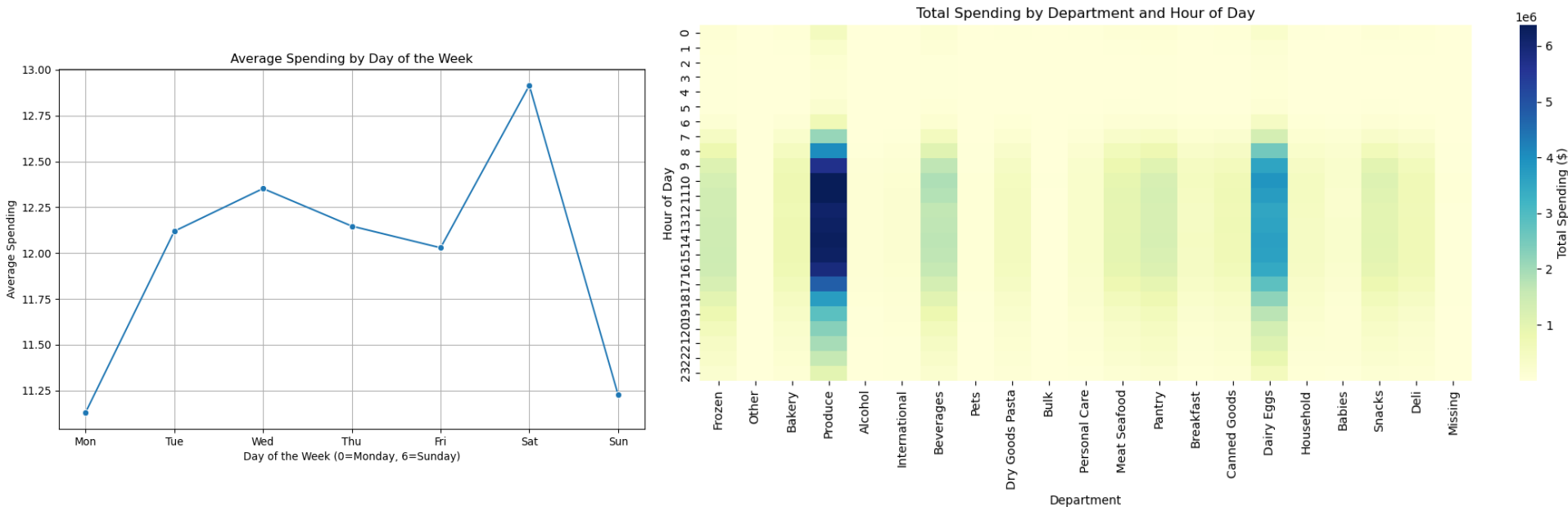
Peak Hour: 10 AM

Order Activity Rises: From 8 AM. **Decline Starts:** After 4 PM

High Frequency Window: 10 AM. **Lowest Activity:** 12 AM – 4 AM

 **Insight:** Focus advertising during low-traffic periods (e.g., early mornings & mid-week) to boost engagement when order volume is low.

Peak Spending Time & Ad Strategy



Peak Spending Time: 10 AM – 4 PM (📌 Especially Saturdays)

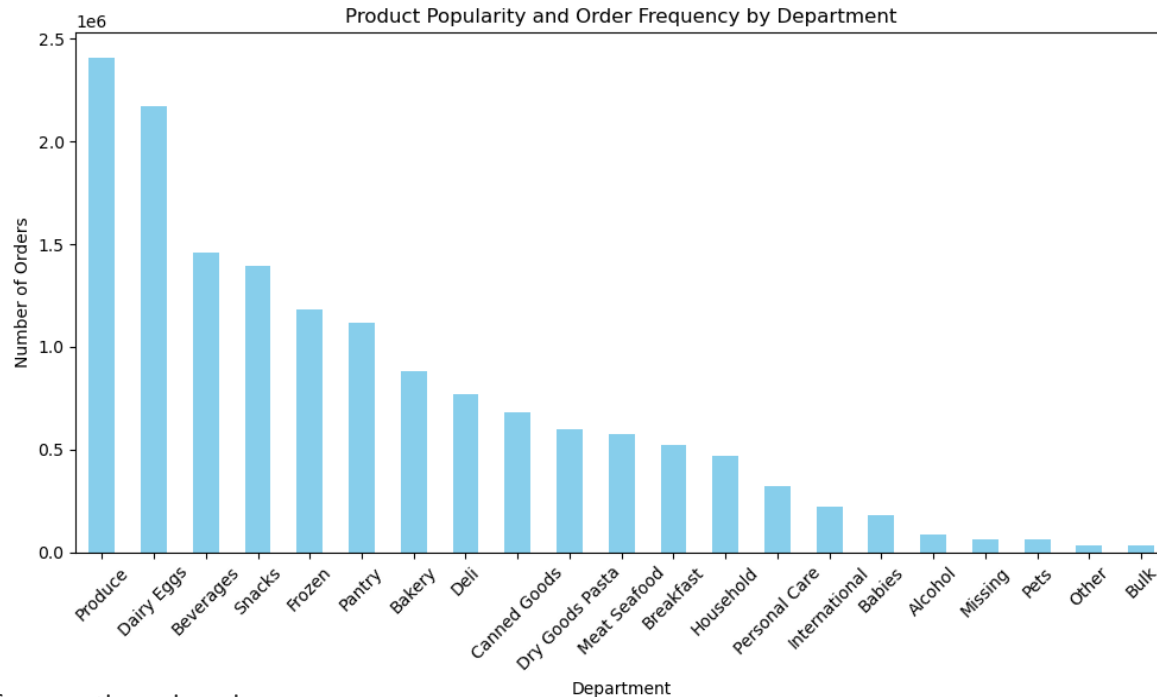


Top Departments: Produce 🥬 & Dairy/Eggs 🥚



Insight: Run ads for high-demand products during peak hours to boost visibility and maximize conversions

Top Product Categories by Order Frequency



1. Produce – Most frequently ordered



2. Dairy & Eggs – Strong consistent demand



3. Beverages – High customer interest

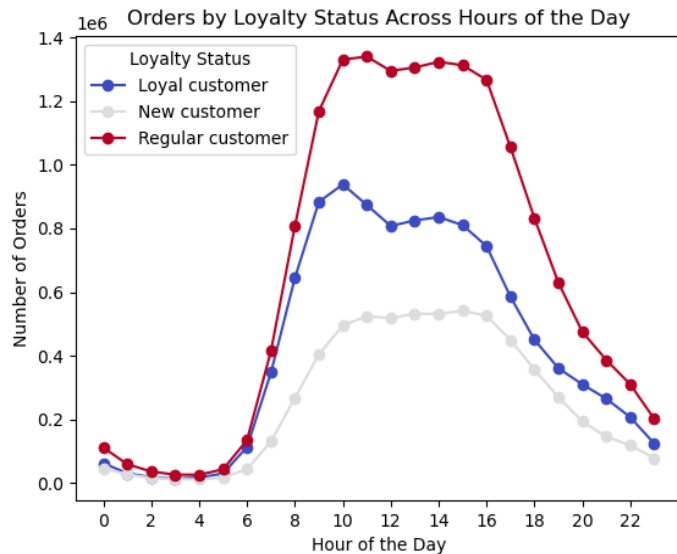


4. Snacks – Popular among frequent buyers

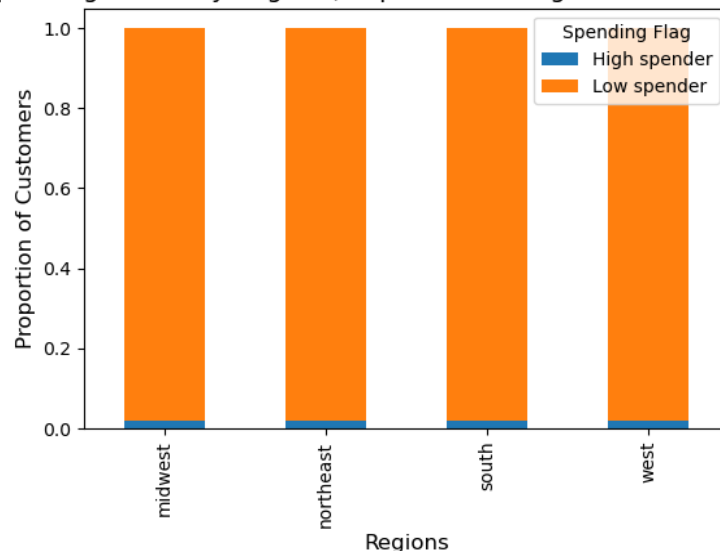


Insight: Focus marketing and inventory on these top departments to align with customer preferences.

Customer Ordering Habits by Loyalty & Region



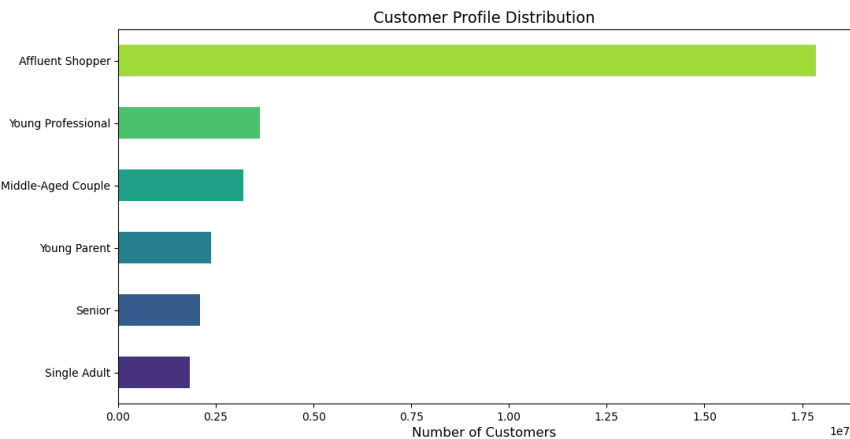
Spending Habits by Region (Proportions of High and Low Spender:



- Regular customers place the most orders, mainly between 10 AM – 4 PM
- Loyal customers follow a similar pattern with fewer orders
- New customers have the lowest and more evenly spread order frequency
- Most customers across all regions are **low spenders** (avg. order < \$10)
- High spenders are rare, with little variation across regions

 **Insight:** Focus strategies on regular customers and explore ways to increase spend per order across all regions.

Demographic Classifications & Customer Spending Trends

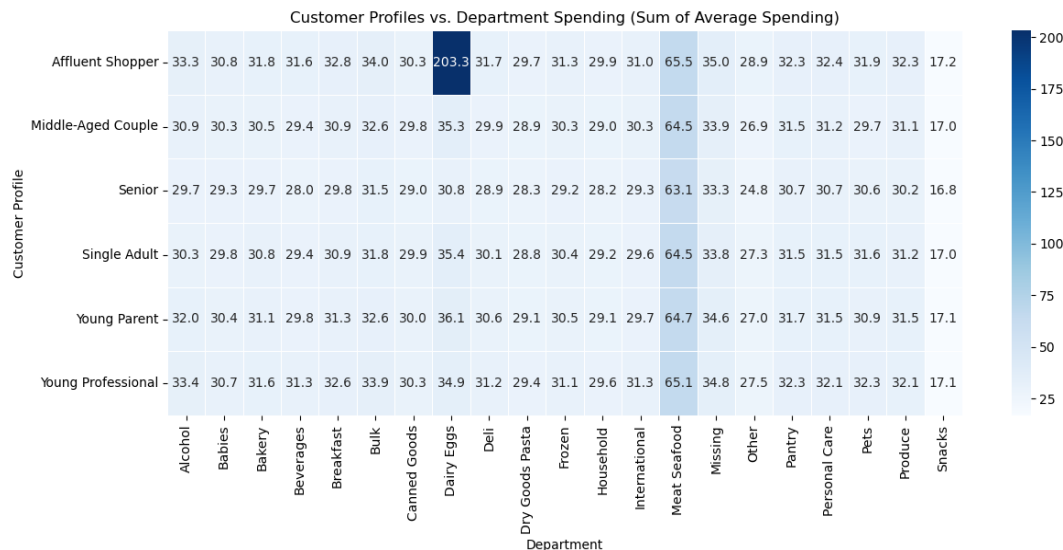


Customer Profiles Based on Demographics

- **Affluent Shopper:** High-income ($\geq \$90,000$)
- **Young Professional:** Aged 25-35 with decent income ($\geq \$40,000$)
- **Senior:** Aged 60+ (regardless of income/dependents)
- **Middle-Aged Couple:** Aged 35-60 with dependents
- **Young Parent:** Under 35 with dependents
- **Single Adult:** Under 60 with no dependents



Insight: Affluent Shoppers spend most, especially on Dairy & Eggs 🧺🥚 — perfect for targeted promos!







Spending Trends by Profile

- **Affluent Shopper:** Highest spending in **Dairy & Eggs** 🧺🥚, followed by **Meat & Seafood** 🍖🐟
- **Other Profiles:** Top spending in **Meat & Seafood** 🍖🐟, followed by **Dairy & Eggs** 🧺🥚


Summary of Findings and Actionable Recommendations


Key Findings:


 Peak orders: Saturday at 10 AM; traffic starts at 8 AM and drops after 4 PM.

 Highest spending: 10 AM–4 PM, especially Saturdays. Focus: Produce  & Dairy & Eggs .


 Top departments: Produce, Dairy & Eggs, Beverages, Snacks.


 Regular customers order the most; new users order less.


 Most customers are low spenders (< \$10/order) across all regions.


 Affluent Shoppers spend the most, especially on Dairy & Eggs.

Recommendations:

 Target Peak Hours: Focus ads on Saturdays, 10 AM–4 PM—promote Produce & Dairy/Eggs.

 Customer Engagement: Reward regulars & loyal users with tailored deals.

 Regional Focus: Boost efforts in the South; engage Northeast with campaigns.

 Profile-Based Offers: Target Affluent Shoppers (Midwest) with Dairy/Eggs; promote Meat & Seafood to Young Pros & Parents.



05

Pig E. Bank



Project Overview



Project Title

**Client Churn Analysis for Pig E.
Bank**




Objective

Identify key factors contributing to customer churn and recommend targeted strategies to improve retention



Business Context

 Pig E. Bank seeks to reduce churn by understanding why clients leave, aiming to boost retention, growth, and customer satisfaction.



Data & Methodology



Data Sources

Pig E. Bank client dataset



Key Questions

- ✓ Which customer demographics are most likely to leave the bank?
- ✓ What behavioural patterns (e.g. inactivity, product usage) predict churn?
- ✓ How does product or service usage influence retention?



Tools & Techniques

Excel – for data cleaning, pivot tables, summary statistics, visualizations, and decision tree logic.



Data Summary: Client Loss Risk Factors

	Count of Customer_ID	Percentage of Customer
Left	204	21%
Stayed	787	79%
Grand Total	991	100%

Estimated_Salary			
	Left	Stayed	Grand Total
0 - 100000	102	401	503
100000 - 200000	102	384	486
N/A		2	2
Grand Total	204	787	991

Count of Balance			
	Left	Stayed	Grand Total
0	56	293	349
1 - 100000	25	130	155
100000 - 200000	120	364	484
200000 - Above	3		3
Grand Total	204	787	991

Count of Age			
	Left	Stayed	Grand Total
Young 18-29	10	139	149
Middle Age 30-59	173	617	790
Senior 60 and more	21	31	52
Grand Total	204	787	991

Count of Active_Membership_Status			
	Left	Stayed	Grand Total
Active	61	442	503
Inactive	143	345	488
Grand Total	204	787	991

Count of Credit_Card_Status			
	Left	Stayed	Grand Total
Has Credit Card	144	556	700
No Credit Card	60	231	291
Grand Total	204	787	991

Count of Num_of_Products			
	Left	Stayed	Grand Total
1	142	368	510
2	32	414	446
3	28	5	33
4	2		2
Grand Total	204	787	991

Count of Tenure			
	Left	Stayed	Grand Total
0	6	32	38
1	30	73	103
2	27	83	110
3	20	81	101
4	17	69	86
5	20	80	100
6	21	73	94
7	15	78	93
8	20	88	108
9	21	87	108
10	7	43	50
Grand Total	204	787	991

Descriptive Statistics

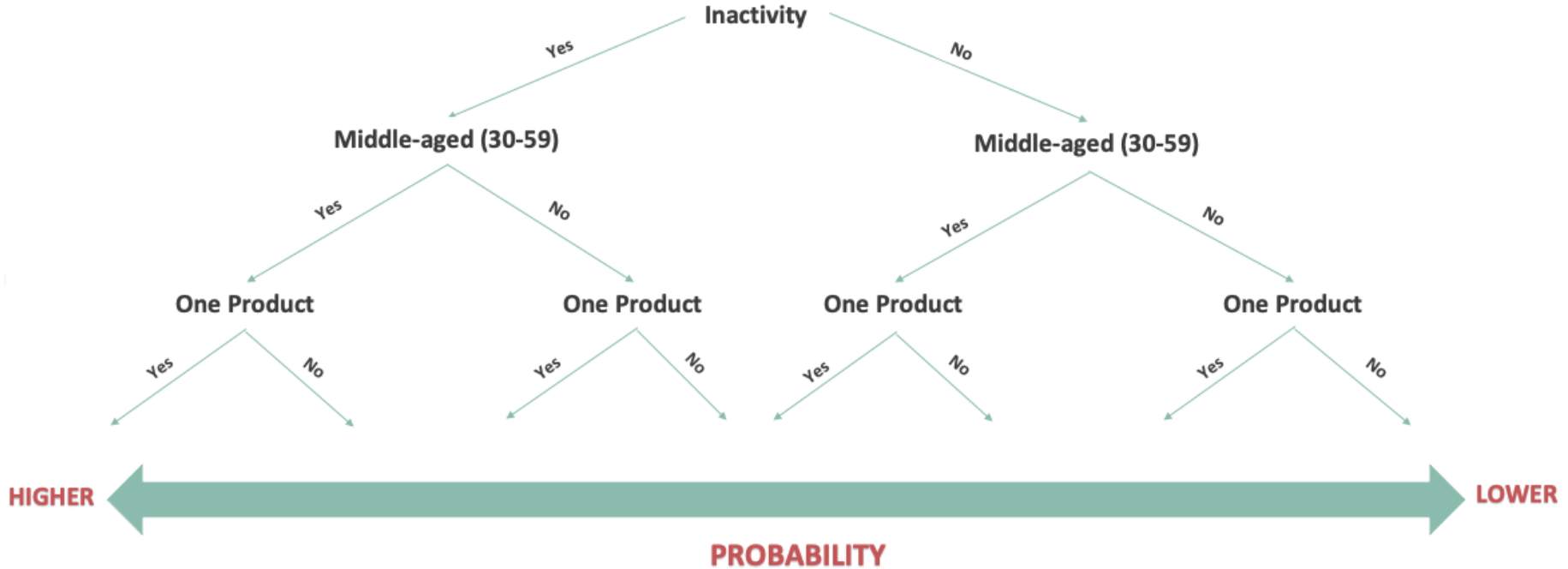
	All customer data					Customers who have closed their accounts			
	Mean	Max	Min	Mode		Mean	Max	Min	Mode
Country				France	Country				France
Gender				Male	Gender				Female
Age	39	82	18		Age	45	69	22	
Credit_Score	648	850	376		Credit_Score	635	850	376	
Tenure	5	10	0		Tenure	5	10	0	
Balance	\$78.002,0	\$213.146,0	\$0,00		Balance	\$90.239,0	\$213.146,0	\$0,00	
Num_of_Products	2	4	1		Num_of_Products	1	4	1	
Credit_Card_Status				Has Credit Card	Credit_Card_Status				Has Credit Card
Active_Membership_Status				Active	Active_Membership_Status				Inactive
Estimated_Salary	\$98.591	\$199.725	\$371		Estimated_Salary	\$97.155	\$199.725	\$417	
Account_Exit_Status				Stayed	Account_Exit_Status				Left

Analysis:

- 1.Age Group:** Middle-aged customers (30-59) account for 85% of churn.
- 2.Balance & Salary:** No clear link between balance or salary and churn.
- 3.Membership:** Inactive customers have higher churn (70%).
- 4.Products:** Fewer products lead to higher churn.
- 5.Credit Card:** Customers without credit cards are more likely to leave.

	Customers Still with the Bank			
	Mean	Max	Min	Mode
Country				France
Gender				Male
Age	37	82	18	
Credit_Score	652	850	411	
Tenure	5	10	0	
Balance	\$74.807,0	\$197.041,0	\$0,00	
Num_of_Products	2	3	1	
Credit_Card_Status				Has Credit Card
Active_Membership_Status				Active
Estimated_Salary	\$98.965	\$199.662	\$371	
Account_Exit_Status				Stayed

Decision Tree





Top Factors Contributing to Client Loss:


- Middle-aged customers (30-59) make up most of the churn.
- Inactive membership is the strongest predictor of loss.
- Customers with only one product are more likely to leave.


Summary of Findings and Actionable Recommendations


Key Findings:

 Middle-aged clients (30–59) make up 85% of those who left.


 Inactive members are the most likely to churn (70% of exits).


 Single-product users have the highest risk of leaving (70% churn).


 No credit card slightly increases churn risk.


 Balance & Salary don't show a clear link to churn.

Recommendations:

 Engage Middle-Aged Clients Offer tailored solutions and personalized support.

 Encourage Active Membership Launch loyalty perks or app engagement reminders.

 Promote Multi-Product Use Introduce bundles and personalized upselling strategies.

 Incentivize Credit Card Adoption Provide signup bonuses or exclusive cardholder benefits.



06

Climate Change & Agriculture



Project Overview



Project Title

**Climate Change Impact on
Agriculture**



Objective

Analyse how environmental factors affect crop yield and economic impact to support sustainable agricultural strategies.



Business Context

📌 With rising climate pressures, agriculture is increasingly vulnerable. This project supports evidence-based policy and investment decisions to enhance food security, sustainability, and climate resilience.



Data & Methodology



Data Sources

Climate Change Impact on Agriculture – Kaggle



Key Questions

- ✓ How has crop yield evolved globally from 1990 to 2024?
- ✓ Which environmental factors (temperature, precipitation, CO₂) most influence crop yield?

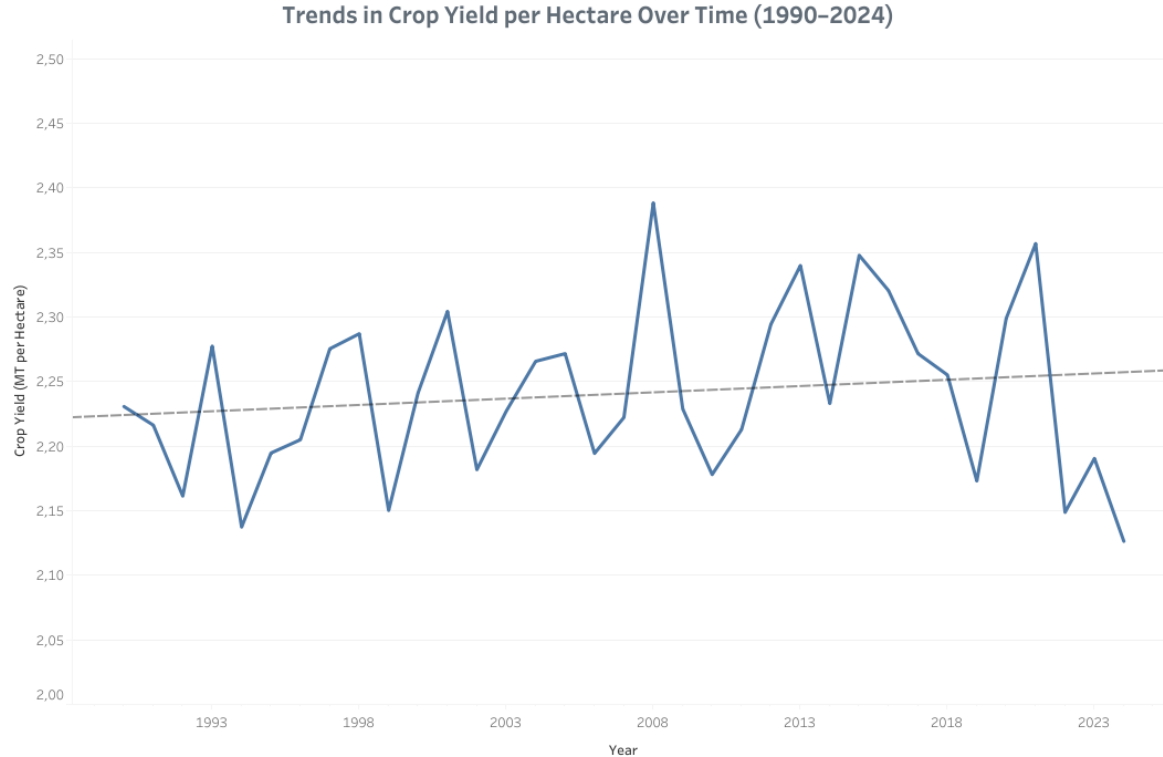


Tools & Techniques

Python and Tableau – used for data cleaning, regression, clustering, correlation analysis, and time-series forecasting.

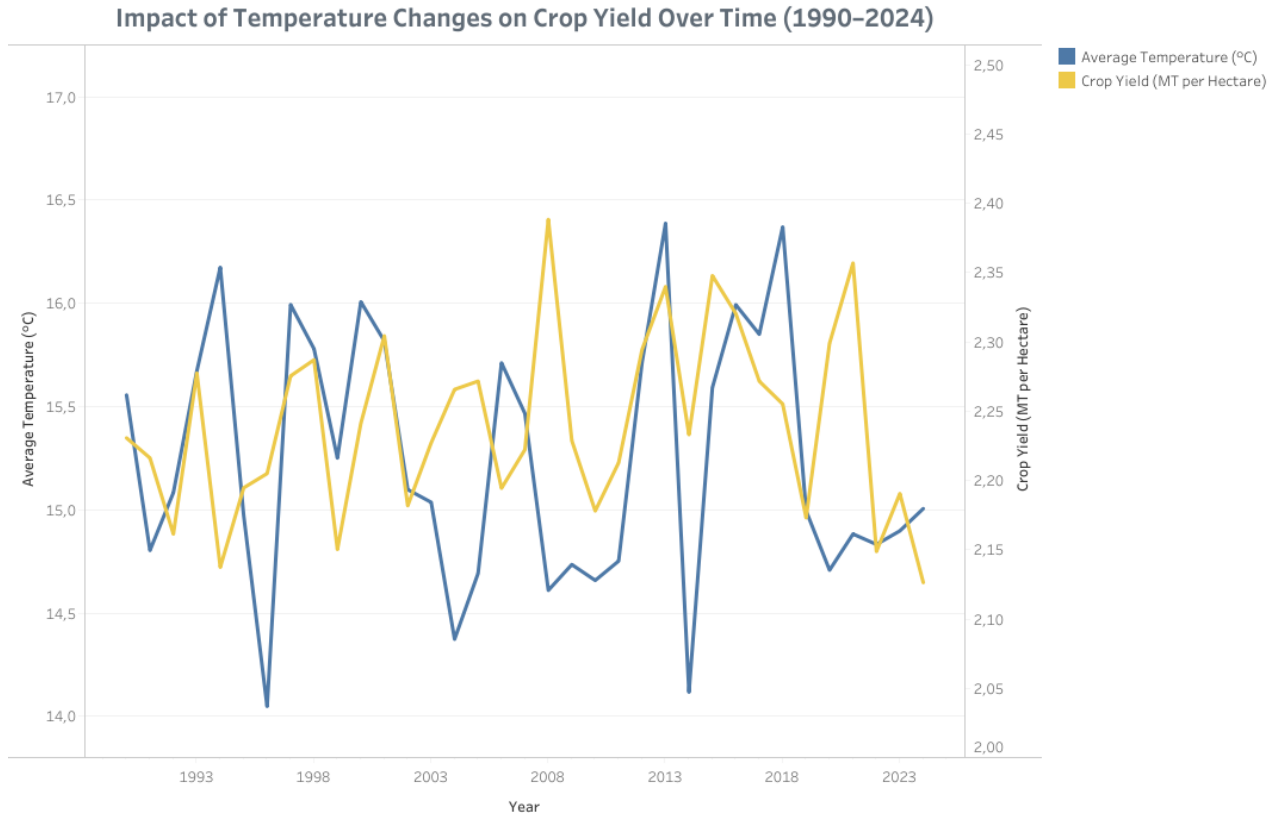


How Has Crop Yield Changed from 1990 to 2024?



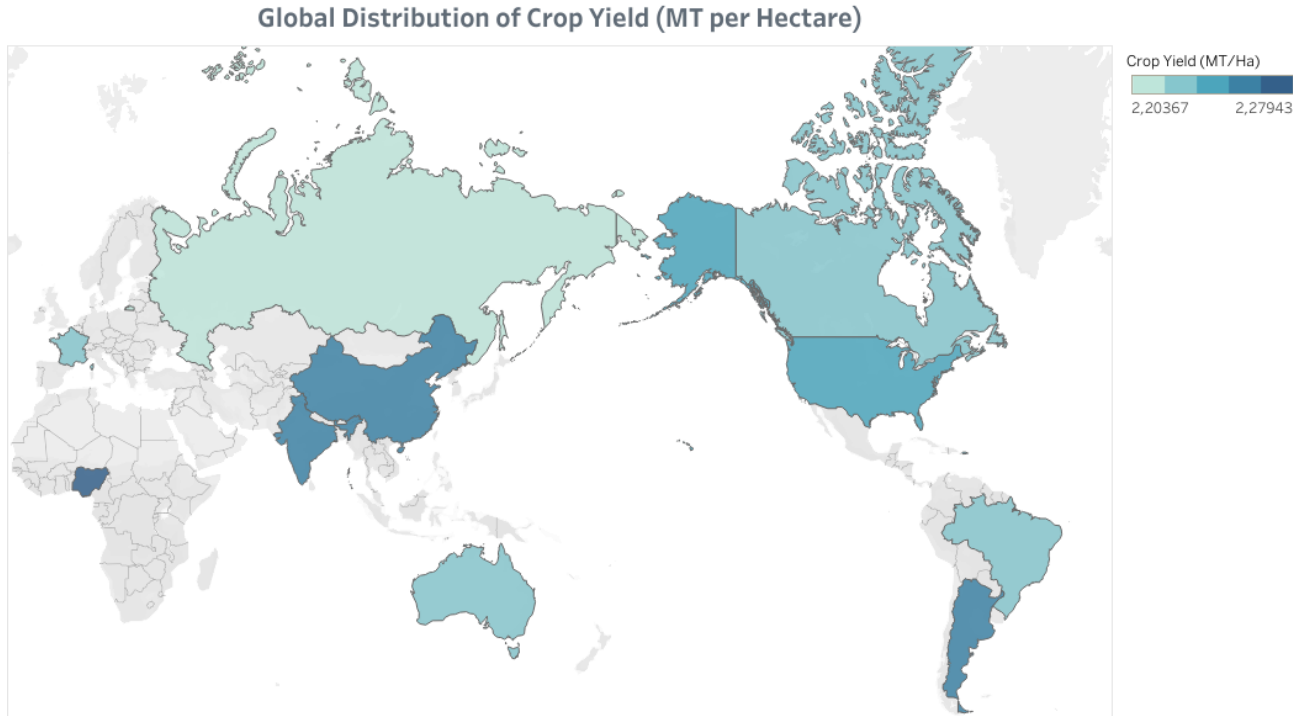
Insight: Global crop yields per hectare have stayed stable, suggesting farming improvements may be balancing climate change effects.


Does Temperature Affect Crop Yield?



Insight: Temperature changes coincide with yield dips in some years, but the relationship is not consistently strong.

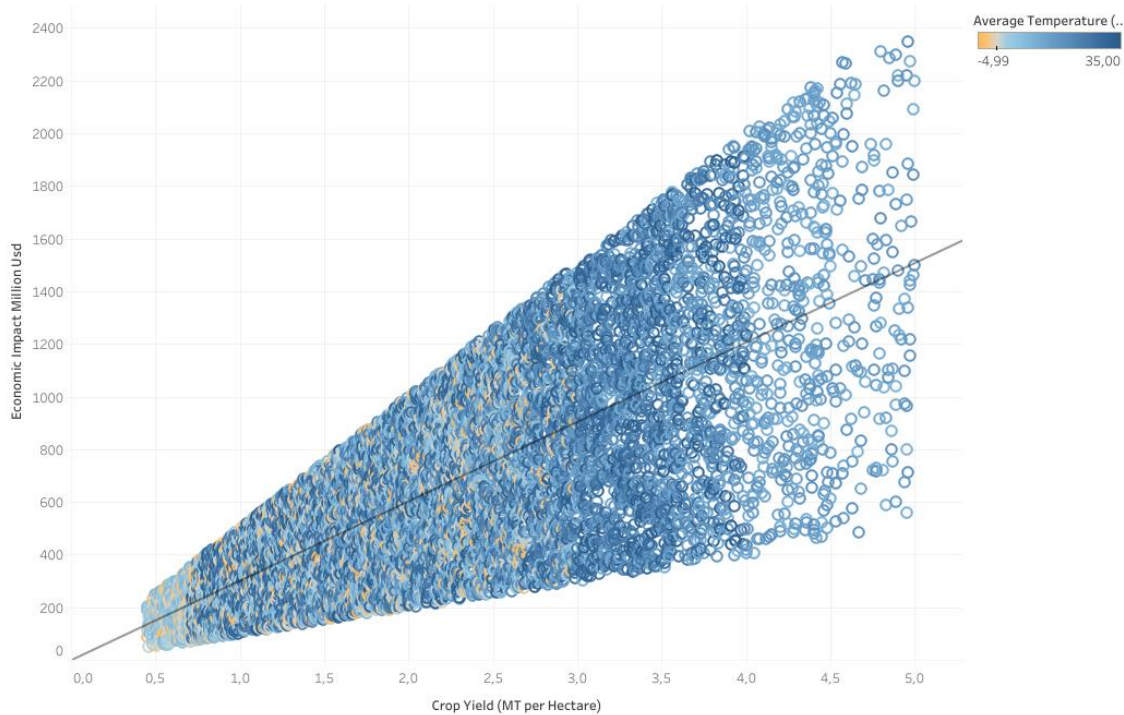
Where Are the Highest Yields?



 **Insight:** Argentina, China, India, and Nigeria lead in crop yields, showing that high yield isn't limited to wealthy nations—environment and farming practices play a key role. Wide yield variation underscores the need for region-specific strategies.

Is Crop Yield a Key Driver of Economic Output?

Correlation Between Crop Yield and Economic Impact of Agriculture and Temperature Variations



Insight: Crop yield strongly drives economic impact, while lower temperatures reduce yields—highlighting the need to protect productivity amid climate change.

Confirming Yield–Economic Impact Link Using Machine Learning & Statistical Tests




Evidence 1: Crop Yield is Stationary

Python code:

```
from statsmodels.tsa.stattools import adfuller
# Perform the Dickey-Fuller test on the crop yield data (after resampling and dropping NaNs)
result = adfuller(df_subset_resampled.dropna())
# Print the test statistic, p-value, and critical values
test_statistic, p_value, used_lag, n_obs, critical_values, icbest = result
# Print the results
print(f"Test Statistic: {test_statistic}")
print(f"P-value: {p_value}")
print("Critical Values:")
for key, value in critical_values.items():
    print(f"    {key}: {value}")
```

•Dickey-Fuller Test Results

- **Test Statistic:** -4.226
- **P-value:** 0.0006  (*Rejects non-stationarity*)

•**Conclusion:** Crop yield data is **stationary**, making it reliable for time-based economic analysis.



Insight Confirmed:

 Higher crop yield is a strong and significant driver of agricultural economic impact.



Evidence 2: Strong Predictive Power

Python code:

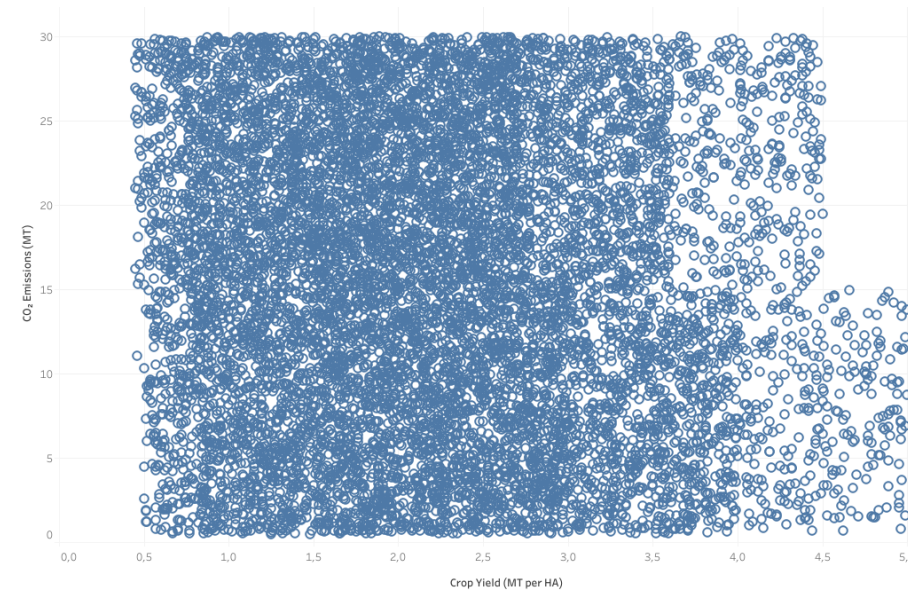
```
# Prepare data for regression
X = df_subset['crop_yield_mt_per_ha']
y = df_subset['economic_impact_million_usd']
# Add a constant to the model (for intercept)
X = sm.add_constant(X)
# Fit the regression model
model = sm.OLS(y, X).fit()
# Get the summary of the regression
model.summary()
```

•Linear Regression Analysis

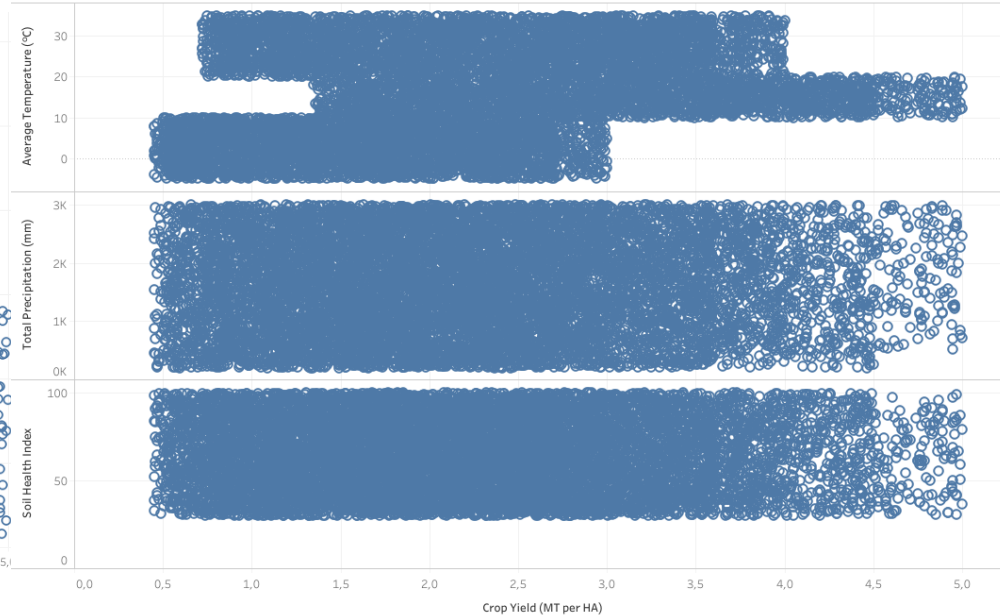
- **$R^2 = 0.528$** → Yield explains 53% of the variation in economic impact.
- **Coefficient:** +\$301.64M per 1 MT/HA increase in yield
- **P-value:** 0.000 → *Highly significant*
- **Durbin-Watson:** 2.025 → *No autocorrelation*

Environment's Broader Role in Agriculture

Relationship Between Crop Yield and CO₂ Emissions



Exploring Environmental Factors That Influence Crop Yield



Insight: No strong correlation observed between crop yield per hectare and CO₂ emissions, temperature, precipitation, or soil health individually, as shown by the scattered data with no clear trend. No Strong individual factor drives yields- climate effects are complex.

Summary of Findings and Actionable Recommendations

Key Findings:



Crop yield drives economic impact in agriculture.



Temperature affects yield, but water and soil health are key.



Top yielders: Argentina, China, India, Nigeria; high yield isn't just for developed countries.



Yield varies regionally, needing tailored strategies.



Stability: Yield per hectare is stable, helping offset climate change.



Economic link: Strong correlation between yield and economic impact.

Recommendations:



Deep-dive into regional or crop-specific analyses for granular insights.



Explore **time-lag effects**, like delayed economic impacts from extreme weather.



Align findings with **SDGs**:

- Goal 2: **Zero Hunger**
- Goal 13: **Climate Action**
- Goal 15: **Life on Land**



Sustainable Agriculture & SDG Recommendations



SDG 2 – Zero Hunger

Target 2.4: Sustainable production & resilience

Finding: High-yield countries are not limited to developed economies—environment and practices matter more.

Action: Support climate-resilient farming, better irrigation, and soil health initiatives

Target 2.5: Genetic diversity in agriculture

Finding: Crop yields vary widely by region.

Action: Promote local seed banks and region-specific crop varieties



SDG 13 – Climate Action

Target 13.1: Resilience to climate hazards

Finding: Lower temperatures reduce yields; climate factors act collectively.

Action: Invest in early warning systems, resilient crop research, and adaptive farming calendars.

Target 13.3: Education on climate adaptation
Finding: No strong individual factor drives yield—climate effects are complex.

Action: Promote climate-smart agriculture training and global knowledge sharing.



SDG 15 – Life on Land

Target 15.3: Combat desertification & restore land

Finding: Soil health alone doesn't strongly predict yield.

Action: Support integrated land management to enhance yield through combined environmental improvements.

Target 15.9: Integrate ecosystem values into planning.

Finding: Productive agriculture depends on practices and environment — not just wealth.

Action: Support policies that balance yield with ecosystem sustainability



Thank You!

Your time and attention are appreciated.
Let's connect!



Dounia.elyou@gmail.com



+49 177 8806434



<https://www.linkedin.com/in/dounia-el-yousoufi-1b60a3181/>



<https://github.com/dounia-elyou>